



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

High-resolution depth profiling using a range-gated CMOS SPAD quanta image sensor

Citation for published version:

Ximing, R, Connolly, P, Halimi, A, Altmann, Y, McLaughlin, S, Gyongy, I, Henderson, R & Buller, G 2018, 'High-resolution depth profiling using a range-gated CMOS SPAD quanta image sensor', *Optics Express*, vol. 26, no. 5, pp. 5541-5557. <https://doi.org/10.1364/OE.26.005541>

Digital Object Identifier (DOI):

[10.1364/OE.26.005541](https://doi.org/10.1364/OE.26.005541)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Optics Express

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



High-resolution depth profiling using a range-gated CMOS SPAD quanta image sensor

XIMING REN,^{1*} PETER W. R. CONNOLLY,¹ ABDERRAHIM HALIMI,¹
YOANN ALTMANN,¹ STEPHEN MCLAUGHLIN,¹ ISTVAN GYONGY,²
ROBERT K. HENDERSON,² AND GERALD S. BULLER¹

¹ School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK

² School of Engineering, Institute for Integrated Micro and Nano Systems, The University of Edinburgh, Edinburgh, EH9 3JL, UK

*Ximing.Ren@hw.ac.uk

Abstract: A CMOS single-photon avalanche diode (SPAD) quanta image sensor is used to reconstruct depth and intensity profiles when operating in a range-gated mode used in conjunction with pulsed laser illumination. By designing the CMOS SPAD array to acquire photons within a pre-determined temporal gate, the need for timing circuitry was avoided and it was therefore possible to have an enhanced fill factor (61% in this case) and a frame rate (100,000 frames per second) that is more difficult to achieve in a SPAD array which uses time-correlated single-photon counting. When coupled with appropriate image reconstruction algorithms, millimeter resolution depth profiles were achieved by iterating through a sequence of temporal delay steps in synchronization with laser illumination pulses. For photon data with high signal-to-noise ratios, depth images with millimeter scale depth uncertainty can be estimated using a standard cross-correlation approach. To enhance the estimation of depth and intensity images in the sparse photon regime, we used a bespoke clustering-based image restoration strategy, taking into account the binomial statistics of the photon data and non-local spatial correlations within the scene. For sparse photon data with total exposure times of 75 ms or less, the bespoke algorithm can reconstruct depth images with millimeter scale depth uncertainty at a stand-off distance of approximately 2 meters. We demonstrate a new approach to single-photon depth and intensity profiling using different target scenes, taking full advantage of the high fill-factor, high frame rate and large array format of this range-gated CMOS SPAD array.

© 2018 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

OCIS codes: (030.5260) Photon counting; (040.0040) Detectors; (040.1240) Arrays; (040.1345) Avalanche photodiodes (APDs); (040.3780) Low light level; (110.6880) Three-dimensional image acquisition.

References and links

1. G. Gariépy, N. Krstajić, R. Henderson, C. Li, R. R. Thomson, G. S. Buller, B. Heshmat, R. Raskar, J. Leach, and D. Faccio, "Single-photon sensitive light-in-flight imaging," *Nat. Commun.* **6**, 6021 (2015).
2. M. Perenzoni, L. Pancheri, and D. Stoppa, "Compact SPAD-based pixel architectures for time-resolved image sensors," *Sensors* **16**, 745 (2016).
3. E. Charbon, "Single-photon imaging in complementary metal oxide semiconductor processes," *Phil. Trans. R. Soc. A.* **372**, 20130100 (2014).
4. E. Charbon and M. W. Fishburn, *Single-Photon Imaging* (Springer, 2011), Chap.7.
5. B. Aull, "Geiger-mode avalanche photodiode arrays integrated to all-digital CMOS circuits," *Sensors* **16**, 495 (2016).
6. I. M. Antolovic, S. Burri, R. A. Hoebe, Y. Maruyama, C. Bruschini, and E. Charbon, "Photon-counting arrays for time-resolved imaging," *Sensors* **16**, 1005 (2016).
7. S. P. Poland, N. Krstajić, J. Monypenny, S. Coelho, D. Tyndall, R. J. Walker, V. Devaughes, J. Richardson, N. Dutton, P. Barber, D. D.-U. Li, K. Suhling, T. Ng, R. K. Henderson, and S. M. Ameer-Beg, "A high speed multifocal multiphoton fluorescence lifetime imaging microscope for live-cell FRET imaging," *Biomed. Opt. Express* **6**, 277-296 (2015).
8. D. Bronzi, F. Villa, S. Tisa, A. Tosi, F. Zappa, D. Durini, S. Weyers, and W. Brockherde, "100 000 frames/s 64×32 single-photon detector array for 2-D imaging and 3-D ranging," *IEEE J. Sel. Top. quantum Electron.* **20**, 354-363 (2014).
9. D. Shin, F. Xu, D. Venkatraman, R. Lussana, F. Villa, F. Zappa, V. K. Goyal, F. N. C. Wong, and J. H. Shapiro, "Photon-efficient imaging with a single-photon camera," *Nat. Commun.* **7**, (2016).

10. N. Krstajić, S. Poland, J. Levitt, R. Walker, A. Erdogan, S. Ameer-Beg, and R. K. Henderson, "0.5 billion events per second time correlated single photon counting using CMOS SPAD arrays," *Opt. Lett.* **40**, 4305-4308 (2015).
11. F. Villa, R. Lussana, D. Bronzi, S. Tisa, A. Tosi, F. Zappa, A. Dalla Mora, D. Contini, D. Durini, S. Weyers, and W. Brockherde, "CMOS imager with 1024 SPADs and TDCS for single-photon timing and 3-D time-of-flight," *IEEE J. Sel. Top. Quantum Electron.* **20**, (2014).
12. A. McCarthy, R. J. Collins, N. J. Krichel, V. Fernández, A. M. Wallace, and G. S. Buller, "Long-range time-of-flight scanning sensor based on high-speed time-correlated single-photon counting," *Appl. Opt.* **48**, 6241 (2009).
13. A. Maccarone, A. McCarthy, X. Ren, R. E. Warburton, A. M. Wallace, J. Moffat, Y. Petillot, and G. S. Buller, "Underwater depth imaging using time-correlated single-photon counting," *Opt. Express* **23**, 33911 (2015).
14. Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Lidar Waveform-Based Analysis of Depth Images Constructed Using Sparse Single-Photon Data," *IEEE Trans. Image Process.* **25**, 1935-1946 (2016).
15. J. M. Pavia, M. Wolf, and E. Charbon, "Measurement and modeling of microlenses fabricated on single-photon avalanche diode arrays for fill factor recovery," *Opt. Express* **22**, 4202-4213 (2014).
16. G. Intermite, A. McCarthy, R. E. Warburton, X. Ren, F. Villa, R. Lussana, A. J. Waddie, M. R. Taghizadeh, A. Tosi, F. Zappa and G. S. Buller, "Fill-factor improvement of Si CMOS single-photon avalanche diode detector arrays by integration of diffractive microlens arrays," *Opt. Express* **23**, 33777-33791 (2015).
17. T. Al Abbas, N. A. W. Dutton, O. Almer, S. Pellegrini, Y. Henrion, and R. K. Henderson, "Backside illuminated SPAD image sensor with 7.83 μm pitch in 3D-Stacked CMOS technology," in *Electron Devices Meeting (IEDM)*, 2016 IEEE International (IEEE, 2016), pp. 1-8.
18. E. R. Fossum, J. Ma, S. Masoodian, L. Anzagira, and R. Zizza, "The quanta image sensor: Every photon counts," *Sensors* **16**, 1260 (2016).
19. N. A. W. Dutton, I. Gyongy, L. Parmesan, S. Gneccchi, N. Calder, B. R. Rae, S. Pellegrini, L. A. Grant, and R. K. Henderson, "A SPAD-based QVGA image sensor for single-photon counting and quanta imaging," *IEEE Trans. Electron Devices* **63**, 189-196 (2016).
20. F. Yang, Y. M. Lu, L. Sbaiz, and M. Vetterli, "Bits from photons: Oversampled image acquisition using binary poisson statistics," *IEEE Trans. image Process.* **21**, 1421-1436 (2012).
21. S. H. Chan, O. A. Elgandy, and X. Wang, "Images from Bits: Non-Iterative Image Reconstruction for Quanta Image Sensors," *Sensors* **16**, 1961 (2016).
22. I. Gyongy, A. Davies, N. A. W. Dutton, R. R. Duncan, C. Rickman, R. K. Henderson, and P. A. Dalgarno, "Smart-aggregation imaging for single molecule localisation with SPAD cameras," *Sci. Rep.* **6**, (2016).
23. L. Mertens, M. Sonleitner, J. Leach, M. Agnew, and M. J. Padgett, "Image reconstruction from photon sparse data," *Sci. Rep.* **7**, (2017).
24. J. Busck and H. Heiselberg, "Gated viewing and high-accuracy three-dimensional laser radar," *Appl. Opt.* **43**, 4705-10 (2004).
25. I. Gyongy, N. Calder, A. Davies, N. A. W. Dutton, R. Duncan, C. Rickman, P. Dalgarno, and R. K. Henderson, "A 256x256, 100kFPS, 61% Fill-factor SPAD Image Sensor for Time-resolved Microscopy Applications," *IEEE Trans. Electron Devices* **65**, 547-557 (2018).
26. N. A. W. Dutton, L. Parmesan, S. Gneccchi, I. Gyongy, N. J. Calder, B. R. Rae, L. A. Grant, and R. K. Henderson, "Oversampled ITOF Imaging Techniques using SPAD-based Quanta Image Sensors," In *Proceedings of the International Image Sensor Workshop*, Vaals, The Netherlands, 8-11 June 2015.
27. A. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller, "Single-photon three-dimensional imaging at up to 10 kilometers range," *Opt. Express* **25**, 11919-11931 (2017).
28. Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Robust Bayesian Target Detection Algorithm for Depth Imaging From Sparse Single-Photon Data," *IEEE Trans. Comp. Imaging* **2**(4), 456-467 (2016).
29. L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D* **60**, 259-268 (1992).
30. A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.* **4**, 490-530 (2005).
31. J. Salmon, Z. Harmany, C.-A. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *J. Math Imaging Vis.* **48**, 279-294 (2014).
32. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.* **16**, 2080-2095 (2007).
33. R. Ammanouil, A. Ferrari, and C. Richard, "A graph Laplacian regularization for hyperspectral data unmixing," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 1637-1641.
34. A. Ng, M. Jordan, and Y. Weiss, "On Spectral Clustering: Analysis and an Algorithm," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2001, pp. 849-856.
35. S. Amari and S. Douglas, "Why natural gradient?" in *Proc. IEEE ICASSP-98*, vol. 2, May 1998, pp. 1213-1216.
36. A. Halimi, C. Mailhes, J.-Y. Tournet and H. Snoussi, "Bayesian Estimation of Smooth Altimetric Parameters: Application to Conventional and Delay/Doppler Altimetry," *IEEE Trans. Geosci. Remote Sens.* **54**(4), 2207-2219 (2015).
37. K. Cui, X. Li, and R. Zhu, "A high-resolution programmable Vernier delay generator based on carry chains in FPGA," *Rev. Sci. Instrum.* **88**, 064703 (2017).

38. A. M. Wallace, J. Ye, N. Krichel, A. McCarthy, R. J. Collins, and G. S. Buller, "Full wave form analysis for long-range 3D imaging laser radar," *EURASIP J. Adv. Signal Process.* **1**, 896708 (2010).
 39. D. Shin, F. Xu, F. N. Wong, J. H. Shapiro, and V. K. Goyal, "Computational multi-depth single-photon imaging," *Opt. Express* **24**(3), 1873-1888 (2016).
 40. W. Wang, J. Yan, N. Xu, Y. Wang, and F-H Hsu, "Real-time high-quality stereo vision system in FPGA," *IEEE Trans. Circuits Syst. Video Technol.* **25**, 1696-1708 (2015).
 41. D. G. Bailey, *Design for Embedded Image Processing on FPGAs* (John Wiley & Sons, 2011), pp. 1-19.
 42. A. Maccarone, A. McCarthy, X. Ren, R.E. Warburton, A.M. Wallace, J. Moffat, Y. Petillot, and G.S. Buller, "Underwater depth imaging using time-correlated single-photon counting," *Opt. Express* **23**, 33911-33926 (2015).
 43. A. Halimi, A. Maccarone, A. McCarthy, S. McLaughlin, and G.S. Buller, "Object depth profile and reflectivity restoration from sparse single-photon data acquired in underwater environments," *IEEE Trans. Comput. Imaging* **3**, 472-484 (2017).
-

1. Introduction

In recent years, advanced silicon complementary metal-oxide-semiconductor (CMOS) single-photon avalanche diode (SPAD) arrays have shown excellent potential for high-resolution, high-speed, high-efficiency and low-noise time-resolved imaging [1, 2] as well as low-light level imaging [3–5]. They have emerged as candidates for solid-state image sensors for single-photon imaging applications, such as fluorescence lifetime imaging microscopy (FLIM) [6, 7] and object depth profiling in photon-starved conditions [8, 9]. At detector and system level, CMOS SPAD image sensors have a significant advantage in cost and read-out noise compared to off-the-shelf single-photon sensitive cameras such as electron-multiplying charge-coupled devices (EMCCDs), intensified CCDs (ICCDs), and scientific CMOS (sCMOS) image sensors. EMCCDs and ICCDs usually require thermoelectric cooling to lower their dark current, and they operate at comparatively low frame rates (<50 frames per second) as CCD circuitry is used to read out signals. Currently, no EMCCDs are equipped with time-resolved modalities, limiting their practical use in single-photon timing applications. In order to perform object depth profiling, continuous-wave and pulsed indirect time-of-flight 3D imaging systems have been demonstrated using SPAD arrays with time-gated photon counters [8]. Pulsed direct time-of-flight 3D imaging based on SPAD arrays with time-stamping photon counters, e.g. on-chip time-to-digital converters (TDCs) has also been achieved [9], however most SPAD arrays with the time-correlated single-photon counting (TCSPC) functionality available today are typically of a 32×32 format or similar [10, 11]. While these arrays provide 2D spatial resolution without the need for a scanning system, their 3D depth profiling performance is limited by their low fill-factor and relatively small format. Scanning systems using individual single-photon detectors can offer excellent performance due to their relatively high efficiency and optimized detectors [12–14]. Typically, with CMOS SPAD detector arrays using TCSPC, a large proportion of an individual pixel is occupied by the on-chip digital circuitry for picosecond resolution time stamping, resulting in a low fill-factor (<5%). CMOS SPAD image sensors equipped with microlens arrays [15, 16] have been shown to improve fill-factor, albeit under limited f-number illumination. Practical alternatives include developments in 3D-stacked CMOS [17] and in-pixel capacitance-based analogue counters followed by analog-to-digital converters [2]. In this work we consider SPAD arrays in a time-gated mode where each frame is outputted as a bit plane, where each bit represents the absence (i.e. 0) or presence (i.e. 1) of a photon in an individual pixel, and may therefore be considered a single-bit quanta image sensor (QIS) [18]. The binary response of each pixel in a QIS array allows these image sensors to operate at very high frame rates with negligible read-out noise. Crucially, this approach results in a less complex pixel design not requiring picosecond resolution time-stamping, allowing small pixel dimensions and high fill-factor without requiring the use of stacked CMOS configurations. Using this approach, pixel pitch sizes of less than $25\ \mu\text{m}$ with fill factors of larger than 20% and frame rates of greater than 10 kHz have been achieved [2, 19].

Intensity reconstruction using the QIS approach has been developed using oversampled binary

frames [20–23]. Based on an active imaging scheme capable of time-gated viewing [24], we report a single-photon depth profiling system using an advanced time-gated, high fill-factor (61%) SPAD image sensor with binary frame response at a high frame rate (up to 100 kfps) [25]. This provides 3D photon count data, which consists of two spatial dimensions and one temporal dimension. To the best of our knowledge, this is the first experimental demonstration of single-photon depth profiling using a range-gated QIS with a direct time-of-flight approach. Our approach achieves significantly improved three-dimensional image reconstruction, in terms of depth resolution, compared to previous depth profiling schemes using a QIS based on indirect time-of-flight measurements using two interleaved time gates [26], which had a depth resolution of approximately 4 cm. The high fill-factor sensor used in our imager does not include TCSPC functionality so in order to reconstruct a depth profile, a sequence of range-gated single-photon 2D images are acquired. This is achieved by iteratively increasing a delay applied to the sensor gate, synchronized with laser pulses directed at the target. Each 2D image was constructed by summing a series of binary frames (in this paper, we consider a range of between 2 and 2000 frames per delay step). Pixel-wise histograms of photon counts against delay steps can then be built up, in this case using a step size of 10 ps. Depth images with millimeter-level depth uncertainty are constructed from these histograms using both a pixel-wise cross-correlation approach and a bespoke clustering-based image restoration algorithm.

2. Imaging set-up

A schematic of the bistatic active imaging system used for gated-viewing measurements is shown in Fig. 1. The pulsed illumination used in this experiment was provided by a pulsed semiconductor laser diode (PicoQuant LDH series with a peak wavelength of 685 nm) combined with two engineered diffusers (ED1-S20 and ED1-C20, Thorlabs). The output pulse duration of the laser diode was ~ 600 ps full width at half maximum when operated in a high-power output mode at a repetition rate of 5 MHz, triggered by a laser diode driver (PicoQuant PDL 800-B). The two diffusers were employed to uniformly illuminate the target. An illumination lens was used to direct the outgoing diffused beam onto the target scene located at a stand-off distance of ~ 2 meters from the imager. An off-the-shelf single lens reflex (SLR) camera lens (Canon EF 50 mm f/1.8 II lens) was used to collect the scattered return photons and image the target onto the SPAD array sensor (a single-bit QIS). Each binary response of photon detection from the single-bit QIS was time gated at the 5 MHz synchronization rate from the pulsed illumination source with a pre-determined delay. The delay consists of a passive delay and step-increment digital delay by a programmable delay generator (Stanford Research Systems DG645). Using a fixed exposure time per frame (0.5 μ s), a consistent number of frames were acquired for each delay step with an acquisition mode of exposure-time tagging frames.

The main parameters of the imaging system can be seen in Table 1. The digital delay generator has a fixed 100 ns delay between external trigger and signal output which, when combined with the maximum 62 ns programmed delay used in the experiment, resulted in a maximum 162 ns dead time. In order to avoid optical pulses arriving during the dead time, a 5 MHz laser repetition rate was chosen. The total exposure time for a single binary frame is dependent upon the global camera shutter of 10 μ s, during which time the laser, of repetition frequency 5 MHz, produces 50 pulses each initiating an acquisition window of 10 ns. To demonstrate this new depth profiling approach at low average laser power levels (~ 420 μ W), a short range of approximately 2 meters was chosen, and a darkened laboratory environment was used for the measurements.

The 256×256 array was fabricated using a 130 nm CMOS process and was composed of 16 μ m pitch pixels designed to achieve a high fill factor of 61% [25]. The image sensor's average dark count rate (DCR) per pixel at room temperature was approximately 14 k counts per second, neglecting "hot pixels" (as defined below), which were discarded prior to image analysis.

The binary-frame data was read out off-chip through a 64-bit wide and 100 MHz digital

Table 1. Summary of the Main System Parameters

System Parameter	Value/Comment
Laser	Pulsed semiconductor diode (PicoQuant LDH685) with an electrical driver (PicoQuant PDL 800-B)
Laser Repetition Rate	5 MHz
Illumination Wavelength	685 nm
Average Output Power	~420 μ W
Laser Pulse Width	~600 ps full width at half maximum
Optical Diffusers	\varnothing 1" square and circle pattern engineered diffusers with 20° divergence angle (ED1-S20 and ED1-C20, Thorlabs)
Illumination lens	\varnothing 2" and $f=75$ mm (AC254-075-B-ML, Thorlabs)
Objective Lens	Canon EF 50 mm f/1.8 II lens
Receiver Aperture	\varnothing 27.8 mm
Range	~2 meters
Image Sensor	CMOS SPAD array with 5 mm \times 5 mm entire array size <ul style="list-style-type: none"> • 256 \times 256 pixels • 61% fill-factor • 16 μm pixel pitch • Up to 100k frames per second
Excess Bias	2.6 V
Detection Efficiency at $\lambda=685$ nm	*SPAD single-photon detection efficiency (SPDE): ~18% †Sensor external quantum efficiency (EQE): ~11%
Gating	10 ns gate duration
External Passive Delay	33 ns by a delay unit (EG&G Ortec Model DB463)
Digital Delay	Using a digital delay generator (Stanford Research Systems DG645): <ul style="list-style-type: none"> • Delay increased iteratively with delay steps of 10 ps • Number of delay steps used: 1501 • Depth range covered: 2.25 meters
Data Acquisition Module	SPAD array combined with a Xilinx-Spartan-6-FPGA-based integration module (Opal Kelly XEM6310-LX150) with 128 MB dynamic RAM and USB 3.0 interface operated using custom software in MATLAB: <ul style="list-style-type: none"> • Exposure time per frame: 0.5 μs • Exposure-time tagging binary frame in a global-shutter exposure mode • Up to 2000 binary frames per delay step

*SPDE is the probability a photon incident on the detector active area will trigger a detectable avalanche current.

†EQE is the probability that a photon will trigger a detectable avalanche current when incident over the entire detector array, i.e. $EQE = SPDE \times \text{fill-factor}$ [25].

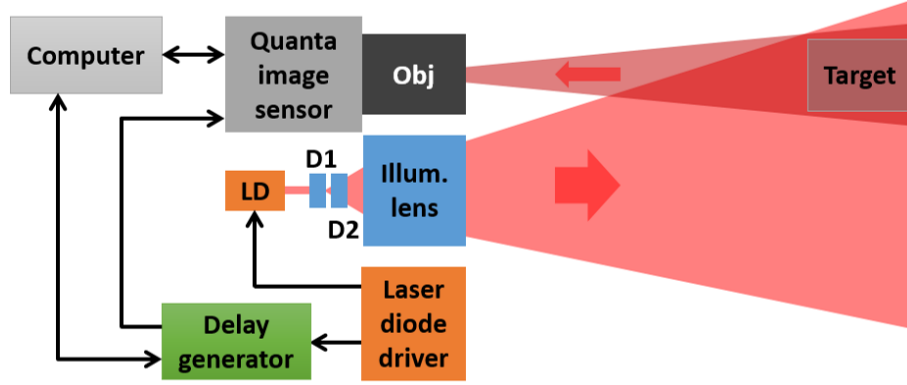


Fig. 1. **Schematic diagram of the imaging system.** The source was a pulsed laser diode (**LD**) with a peak wavelength of 685 nm. Two diffusers (**D1** and **D2**) were placed at the laser output to provide uniform illumination of the target. A 75 mm focal length lens was used to illuminate the target scene located at a stand-off distance of ~ 2 meters from the imager. A single lens reflex (50 mm focal length) camera lens was used as the system objective (**Obj**) to collect the scattered return photons from the target.

output bus under a global-shutter exposure mode [25]. A global gate signal was produced by an on-chip programmable time-gate generator and then imposed on the global shutter. The gate signal was triggered by an external signal from the delay generator which was synchronized to the pulsed laser source. The use of a gate signal with a duration of 10 ns was found to achieve an optimized trade-off between photon detection efficiency and dark count rate of the image sensor. A field-programmable gate array (FPGA) board (Opal Kelly XEM6310-LX150) capable of continuous data streaming at rates of >100 MB/s was used to configure the detector chip and control the gate and transmit signals for the acquisition of image frames. To temporally scan the gate through the entire target depth, the digital delay generator was programmed to iteratively increase the delay in steps of 10 ps, for 1501 steps, with 2000 individual binary frames recorded at each delay step.

3. Creating images and histograms from time-gated oversampled binary frames

As shown in Fig. 2(a), a single binary frame corresponds to the two spatial dimensions (i.e. x and y) of the image sensor's field of view. A number of frames, N_f , are collected for each temporal location (i.e. delay step) and summed to create a grayscale image composed of N_f binary frames (examples shown in Fig. 2(b)).

The summation of photon counts over 256×256 pixels of delay-step-wise grayscale image can be used to build up a histogram of total photon counts versus delay steps with a delay interval of 10 ps, as shown in Fig. 2(b). This represents a temporal sampling approach along the target scene's depth axis. The absence of photon counts scattered from the target (a life-sized polystyrene head in this case, shown in inset of Fig. 2(b)) within the gate corresponds to only background counts. As the gate advances, increasingly more scattered photons return from the target until the entire target scene is captured on the plateau of the gate. As the gate progresses through the target the profile of the scene returns back towards purely background counts once more. This process is displayed as grayscale images at several delay positions in Fig. 2(b). 2000 binary frames per delay step were acquired in order to ensure high-quality ground truth images. However, this delay-step-wise grayscale image can be formed from fewer binary frames, since each frame is time-tagged and a collection of frames can be post-selected after measurement. We

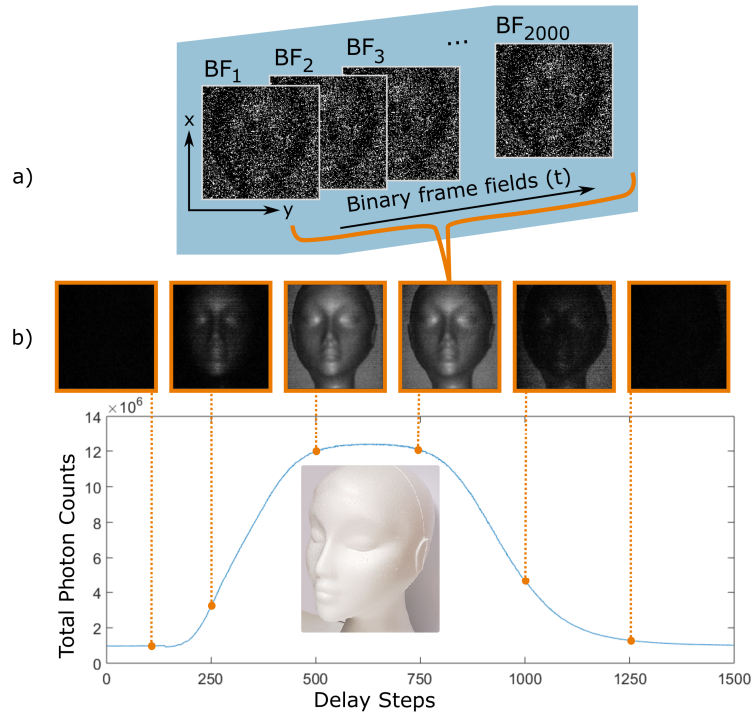


Fig. 2. (a) Examples of individual binary frames (BFs) which are summed to create an image for a single delay step. (b) Examples of delay-step-wise 256×256 grayscale images formed from the summation of 2000 BFs. The approximate position of each example image is shown on the timing histogram. Note that the close-up photograph inset is a side view of the target scene.

examined images formed from a range of binary frames in order to examine the trade-off between image signal-to-noise ratio and measurement time. The pixel-wise histograms of photon counts versus delay steps can be seen in Fig. 3, which illustrates differences in amplitude (i.e. photon counts), gradient and location of leading and falling edges between different example pixels imaged on a flat uniform reference plane. The variation of detector efficiency is likely due to differences in the breakdown voltage between individual detectors. In our depth profiling results, the leading edge mismatches observed can be attributed to the transit time of the gate signal across the array, whilst the falling edge mismatches can be explained by self-heating affecting the SPAD dead time [25].

The variations shown in Fig. 3 illustrate the requirement to acquire a full-field system calibration on a flat uniform reference plane prior to processing of target data. This means that prior to the measurements of a target, the acquisition of 256×256 instrumental response functions (IRFs) was made to determine the response of each individual pixel to a flat uniform surface. Subsequently, the image processing approaches described below were used to determine the depth deviation of the target structure from the flat reference plane.

4. Hot pixel identification and removal

A “hot pixel” is one which has a significantly higher DCR than average, rendering the pixel unusable in many applications [22]. In this work, these hot pixels were removed from the data sets, and the missing pixels were subsequently reconstructed using information contained in the surrounding pixels. For the detector used in this paper we define a hot pixel as one which produces

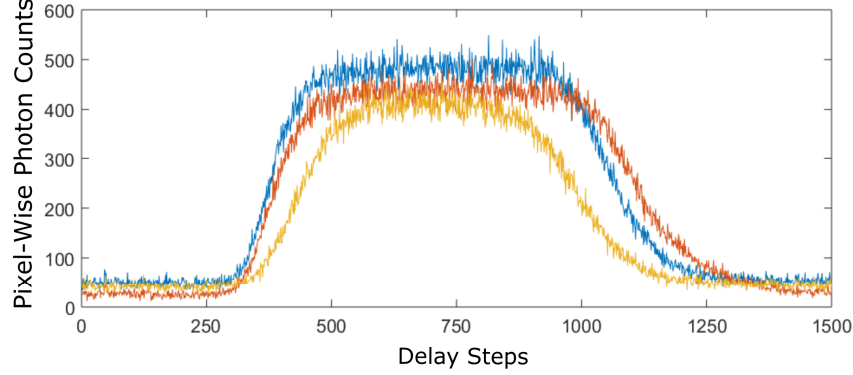


Fig. 3. Histograms of photon counts versus time delay steps for three example pixels from the same target region, illustrating the differences in detector response.

a dark count level higher than 3% of the number of binary frames. This threshold was chosen as it was found to discard the noisiest pixels while preserving a large enough sample to obtain high quality depth reconstruction using the pixel-wise cross-correlation approach described below. A higher threshold (i.e. fewer pixels discarded) resulted in a number of noisy pixels remaining in the reconstructed image, while a lower threshold did not significantly improve the image. This threshold can be defined by the inequality:

$$\frac{\sum_{t=T_0}^{T_1} y_{n,t}^{(d)}}{T_1 - T_0 + 1} > N_f p_d \quad (1)$$

where t denotes the delay step, T_0 and T_1 defining the initial and final delay step of the measurement, $y_{n,t}^{(d)}$ is the number of dark counts recorded by the n th pixel in delay step t , N_f is the number of binary frames used in each timing window and p_d is the fractional threshold of dark counts above which a pixel is considered hot. Here we consider $T_0 = 1$ and $T_1 = 1501$ and use $p_d = 0.03$ which, for the SPAD array used in this work, represented 14,469 pixels, or 22.08% of the entire 256×256 array.

5. Computational methods and reconstruction results

5.1. Pixel-wise cross-correlation

A pixel-wise cross-correlation algorithm was used to obtain a depth estimate for each remaining pixel. Cross-correlation is commonly used in single-photon 3D image reconstruction [12–14, 27, 28] as a simple and fast approach to ascertain depth estimates assuming a reference measurement is present. The spatial offset at each pixel location, n , between the target response and the response of a flat plane (i.e. the IRF denoted by $g_{n,t}$ for the n^{th} pixel) is found by maximizing the cross-correlation between the two functions, described mathematically as:

$$y_{n,t} \star g_{n,t} = \sum_{m=1}^T y_{n,m+t} g_{n,m}^* \quad (2)$$

where \star denotes the cross-correlation between the two functions, $y_{n,m+t}$ denotes the number of counts of the target data, and $g_{n,m}^*$ is the complex conjugate of the IRF at the n^{th} pixel, where

$g_{n,m}^* = g_{n,m}$ as $g_{n,m}$ is real, t is the time bin corresponding to each delay step, with T being the number of delay steps. The resulting function of t is maximized to obtain T_{max} which corresponds to the temporal offset, and therefore the distance, between the flat plane used to record the IRF and the target at that point. The intensity of each pixel is calculated by averaging the values of 100 delay steps from the plateau in the center of the histogram, as shown in Fig. 4.

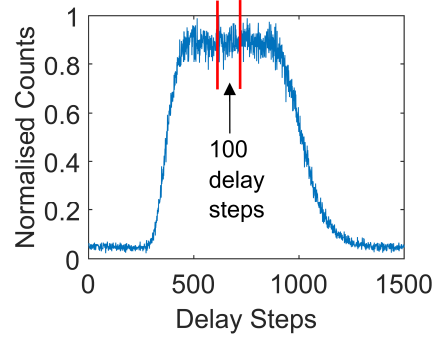


Fig. 4. Timing histogram of photon returns for a single pixel. The counts from 100 delay steps from the plateau representing returns from the target are averaged to provide a value for target intensity in that pixel.

Depth profiles were obtained using this methodology for two targets: a life-sized polystyrene head (approximately $285 \text{ mm} \times 150 \text{ mm} \times 180 \text{ mm}$ in $H \times W \times D$), used for its highly scattering surface and recognizable shape; and a still life arrangement comprising assorted fruit (approximately $180 \text{ mm} \times 180 \text{ mm} \times 200 \text{ mm}$ in $H \times W \times D$) to provide response from a variety of different surface textures and reflectivities. Both target scenes were placed at a distance of approximately 2 meters from the receiver. The detector field of view at the target was approximately $190 \text{ mm} \times 190 \text{ mm}$ square, indicated by the dashed red line overlaying the images in Fig. 5.

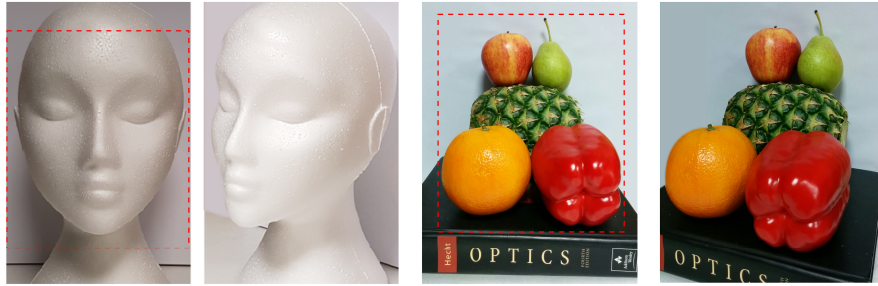


Fig. 5. Images of the two targets used in this work. The first two photographs show the life-sized polystyrene head. The second two photographs show the display of assorted fruit comprised of a pineapple, orange, red pepper, apple and pear. The detector field of view is shown, for both targets, by the red dashed lines (approximately $190 \text{ mm} \times 190 \text{ mm}$). Both targets were placed in front of a flat uniform board to act as a back plane.

Once the hot pixel mask is applied, the missing pixels are reconstructed using a 3×3 median filter approach. For this, the average value of the 8 pixels surrounding the missing pixel is used as the replacement count value in the "hot pixel" position. If there are other hot pixels in the

8 pixel group, they are not used in the averaging. The raw results of the cross-correlation for the polystyrene head and the fruit arrangement are shown in Fig. 6, alongside the corrected cross-correlation which displays the results of the reconstruction with the median filter applied. In both cases 2000 binary frames are used at each delay step.

In addition, we applied this cross-correlation approach to provide depth values from a flat uniform plane in order to quantitatively evaluate the depth uncertainties for the different numbers of binary frames at various illumination power levels. We present the illumination power level in terms of the average photon number per pixel per binary frame (PN_{BF}) for all delay steps used. The depth residuals, which are expressed as one standard deviation from the mean, were recorded for the different numbers of binary frames to produce the results shown in Fig. 7. As the number of binary frames or illumination power increases, lower depth uncertainties are found due to enhanced signal-to-noise ratios. For high numbers of binary frames (i.e. ≥ 500) or high illumination power (i.e. $PN_{BF} \geq 0.51$), as shown in Fig. 7, millimeter scale depth uncertainties have been achieved. These results demonstrate that benefits can be gained through the use of either a greater number of binary frames (i.e. longer acquisition time) or an increased optical power depending upon the application.

The cross-correlation algorithm is imperfect because it does not consider noise from either background light or photon statistics which are more significant at low photon numbers and can lead to errors in peak identification. As expected, at lower numbers of binary frames the quality of the depth reconstruction degrades significantly as the number of collected photons reduces. With the data acquired in these experiments, the cross-correlation required both leading and falling edges of the histogram, meaning more delay steps must be recorded to use the cross-correlation approach than may be necessary. For these reasons, a more advanced algorithm was developed which was able to extract improved depth information from sparse photon data.

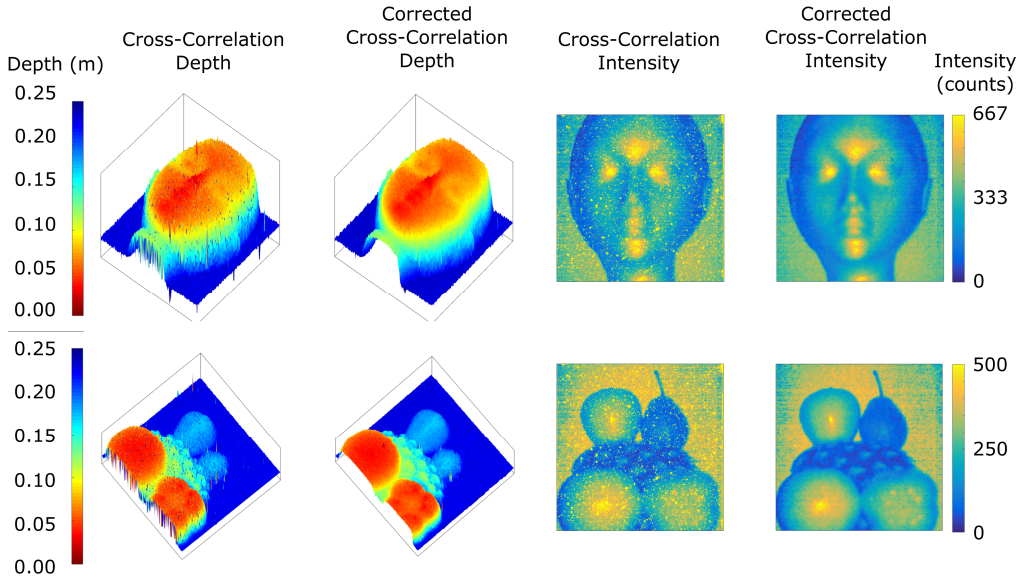


Fig. 6. Depth (left) and intensity (right) maps of polystyrene head (top) and fruit arrangement (bottom) using cross-correlation and the median filter corrected cross-correlation. The intensity color maps belong to the intervals $[0, N_f/3]$ and $[0, N_f/4]$ for the head and fruit respectively, and $N_f = 2000$ in both cases.

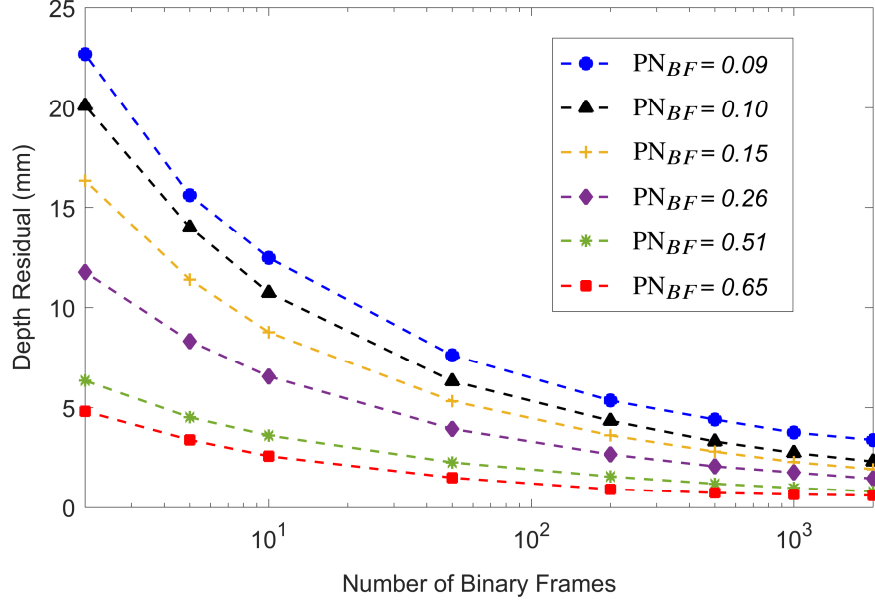


Fig. 7. Standard deviation of the mean depth (i.e. depth residual) as a function of number of binary frames using cross-correlation at various illumination power levels. The calculation of standard deviation was performed on a flat reference plane. The illumination power level is presented in terms of the average photon number per pixel per binary frame (PN_{BF}) for all delay steps used.

5.2. Clustering-based image restoration

In order to improve the quality of the estimated depth and intensity images, a more sophisticated approach should be considered which takes into account the binomial statistics of the data and the spatial correlations in the observed scene. Several strategies have previously been proposed to include spatial correlation and we distinguish between local [29] and non-local [30–32] approaches. The latter will be considered in this paper as they have shown promising results during recent years (see Fig. 8 to illustrate non-local classification). As previously explained, the data is constructed by summing N_f binary images for each temporal delay. This means that the photon counts $y_{n,t}$ of each pixel n , at each delay t , can be assumed to follow a binomial distribution $\mathcal{B}\left(N_f, \frac{s_{n,t}}{N_f}\right)$, where $s_{n,t}$ denotes the average photon counts whose shape is related to the system impulse response. In this paper, we only consider the delays corresponding to the leading edge of the measurements (unlike the cross-correlation approach described earlier, which requires both leading and falling edges of the histogram) and approximates the obtained waveform with an analytical function given by:

$$s_{n,t} = \frac{r_n}{2} \left[1 + \operatorname{erf}\left(\frac{t - d_n}{h_n}\right) \right] + b_n \quad (3)$$

where $\operatorname{erf}(\dots)$ denotes the error function. This model is expressed according to different parameters that include the target's depth profile $d_n \geq 0$, the target's reflectivity of $r_n \geq 0$, the background noise $b_n \geq 0$, and $h_n \geq 0$ which is related to the local orientation of the observed target's surface. Examples of raw data from some pixels which are fitted by the proposed model are shown in Fig. 9.

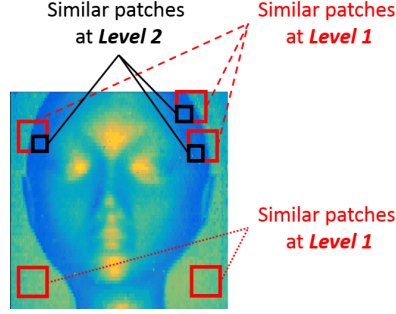


Fig. 8. Illustrative example of non-local classification used in the restorative clustering-based algorithm described in the text. Similar sized patches are used at each level, with subsequent levels containing progressively smaller patches.

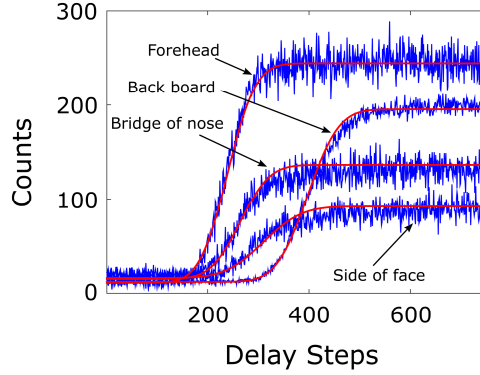


Fig. 9. Examples of photon return data from four pixels from different parts of the target (shown in blue) and their fitting by the analytical error function (in red) for $N_f=1000$ frames.

To improve the quality of these parameters, this paper considers a two step algorithm which can be summarized by considering a clustering stage and an estimation stage:

- **Clustering:** the first step performs a spatial segmentation of the data into homogeneous regions that share similar estimated parameters (e.g. depth, intensity), the goal being to separate the large data sets into small regions that can be processed independently which reduces the computational cost. For this, the proposed method builds on the work [33] that evaluated a graph of similarity for the image pixels, then used it to perform a clustering step by considering the method described in [34]. A direct application of this approach can be done by considering the graph-nodes as the pixels' temporal responses, and their similarities to form the graph's edges. However, due to the high dimension of the considered images, this will lead to a large amount of data requiring costly computational resources (e.g. an image of size 256×256 will lead to a graph of size 65536×65536). To account for the large dimension of the data and reduce the computational cost, a multi-level approach is considered where the graph-nodes are associated with patches of histograms and the cross-correlation estimated parameters associated with these patches. Each level contains patches of a consistent size, while subsequent levels use progressively smaller patches (i.e. finer resolution) as illustrated in Fig. 8. A clustering is then performed using the method described in [34] on the resulting graph leading to smaller sub-graphs, that are

computationally more attractive. Repeating this procedure for ℓ levels will finally provide K clusters and K small similarity sub-graphs \mathbf{W}_k associated with each cluster, K being the total number of clusters over the ℓ levels.

- Estimation: the pixels of each spatial region k are used to estimate the parameters of interest Θ : target's depth, intensity and local orientation and the background noise. This is achieved by considering a Bayesian approach that accounts for the data binomial statistics (i.e. likelihood) and the non-local spatial correlation between the parameters as prior information. More precisely, each sub-graph \mathbf{W}_k is used to construct a Laplacian \mathbf{L}_k that is used to define an inverse covariance matrix for a Gaussian Markov-random-field prior distribution for the parameters of interest Θ . This will lead to the following cost function (i.e. negative log-posterior) for each spatial class

$$C(\Theta) = \mathcal{L}(\Theta) + \phi(\Theta) \quad (4)$$

where \mathcal{L} is the negative log-likelihood function depending on the binomial statistics of the data and $\phi(\Theta)$ is the regularization term that introduces the non-local spatial correlation between the target's estimated parameters. The maximum a-posteriori estimator is then approximated by minimizing this cost function with respect to Θ using a Fisher scoring gradient algorithm that allows a fast convergence, as already reported in [35, 36].

This method has been applied to restore the polystyrene head estimated parameters when reducing the number of considered frames. In such a case, the signal-to-noise ratio decreases, which affects the performance of the clustering step of the proposed algorithm. Therefore, in addition to the hot pixels described in section 4, we remove additional noisy pixels and replace them using a median filter approach described in section 5.1. These additional noisy pixels were defined as those which produce, on average, counts over 30% of saturation over the entire leading edge of the histogram. This resulted in a maximum additional proportion of removed pixels of 13% in the extreme case of minimum frames used per delay step. For consistency in the comparison of the two algorithms, exactly the same corrected data is used for both the clustering-based and cross-correlation algorithms in the rest of this paper. Fig. 10 shows the estimated depth when reducing the number of frames from 2000 to 2 binary frames. This figure clearly shows the benefit of the proposed approach that improves the quality of the estimated depths. Similar improvement is also achieved for the intensity as illustrated in Fig. 11.

In addition to depth and intensity, the proposed approach allows the estimation of other useful parameters including the local orientation of the observed target's surface as shown in Fig. 12. Precisely, the angle between the incident beam direction and the local normal to the surface (i.e. incident angle) is related to the width of the leading edge of the histogram of each pixel, as can be seen in Fig. 9, where the larger width of the leading edge, the higher incident angle. The local orientation map in Fig. 12 shows that higher incident angles occurred at the face's edge and on the nose, as expected. The obtained results have also been evaluated quantitatively by considering the signal-to-reconstruction error (SRE) ratio of the depth and intensity images, where the performances are better for higher SRE. This criterion is given by Equation 5:

$$\text{SRE} = 10 \log_{10} \left(\frac{\|\mathbf{x}\|^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|^2} \right) \quad (5)$$

Here, \mathbf{x} is the reference depth or intensity image (defined as the restored images for 2000 frames), $\hat{\mathbf{x}}$ is the restored image and $\|\mathbf{x}\|^2$ denotes the ℓ_2 norm given by $\mathbf{x}^T \mathbf{x}$. To provide a fair comparison between algorithms, the SRE has been evaluated while discarding hot pixels. The obtained curves are displayed in Fig. 13 showing a clear improvement of the proposed approach with respect to the standard cross-correlation approach.

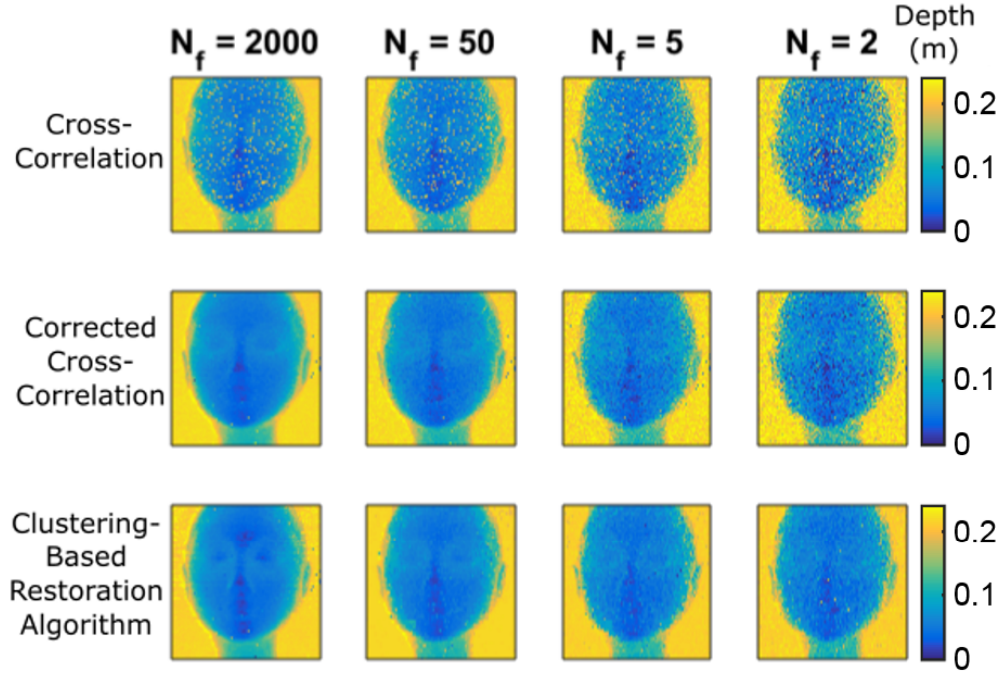


Fig. 10. Depth maps of the polystyrene head for different numbers of binary frames obtained using: top - cross-correlation; middle - corrected cross-correlation; bottom - clustering-based restoration algorithm.

Finally, the average photon number per pixel over all binary frames (PN_{Σ}) for all delay steps used, neglecting hot pixels, and the total exposure time (t) required by each reconstruction method are compared in Table 2, showing fewer photons are required by the proposed algorithm (36% fewer on average) and half as long a total exposure time. The total acquisition time is also provided, showing a limitation of ~ 83 seconds caused by the communication with the external delay generator required for each delay update. This limitation will be significantly reduced by integrating a delay generating functionality directly onto the FPGA, if a picosecond delay step could be achieved [37]. The exposure time could also be reduced further by decreasing the number of delay steps both before and after the gate has reached the target. However, this may slightly lower the quality of the error function fitting, which will reduce the quality of the estimated parameters.

6. Discussion and conclusion

We have constructed a single-photon depth profiling system using a gated CMOS SPAD based binary camera - a single-bit quanta image sensor - with high fill-factor, high frame rate and negligible read-out noise. The detector array does not have TCSPC functionality, but depth resolution of a few millimeters was demonstrated by using a sliding detector gate configuration. The SPAD array was synchronized to pulsed laser illumination and the delay applied to the sensor gate was iteratively increased in 10 ps increments. Thus, a sequence of range-gated images are recorded, and this information is then used to reconstruct the depth and intensity profiles of the target. Each range-gated image can be formed over a series of binary frames.

A new clustering-based image restoration method was shown to better reconstruct depth and

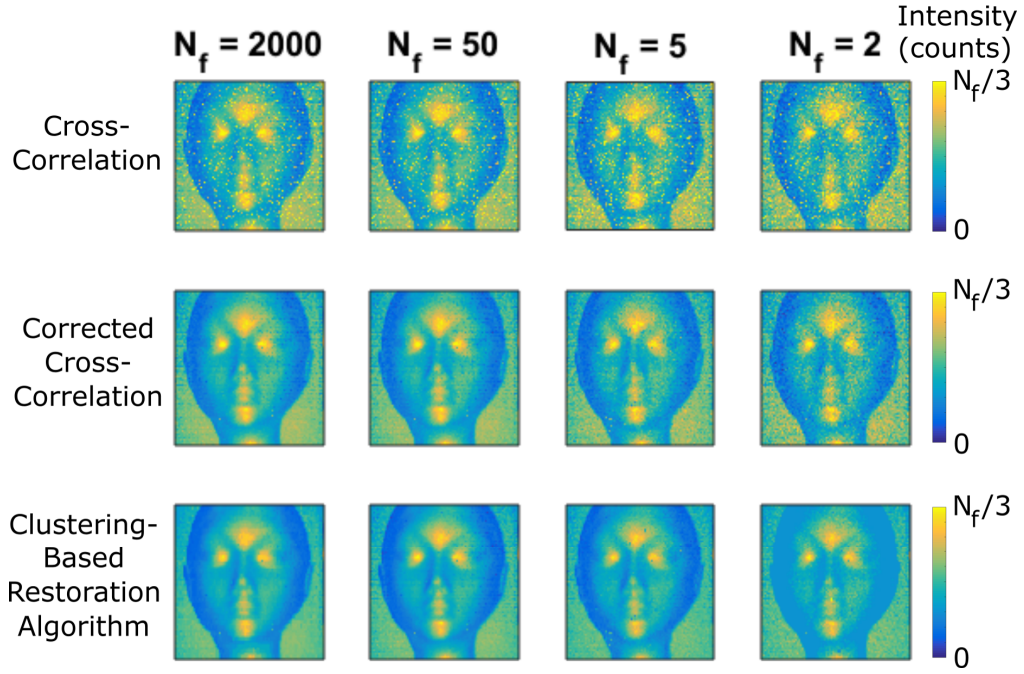


Fig. 11. Intensity maps of the polystyrene head for different numbers of binary frames obtained using: top - cross-correlation; middle - corrected cross-correlation; bottom - clustering-based restoration algorithm. The intensity color maps belong to the interval $[0, N_f/3]$.

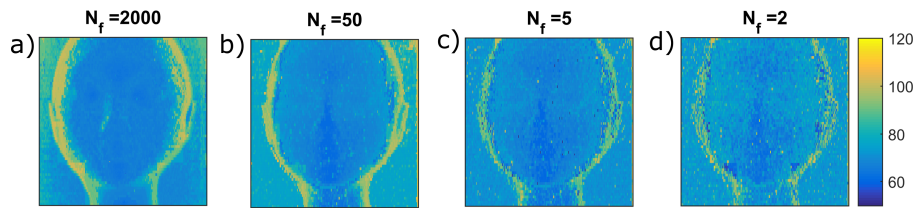


Fig. 12. Local orientation maps obtained with the clustering-based restoration algorithm for the polystyrene head target. Local orientation maps are shown for data sets containing: a) 2000; b) 50; c) 5; d) 2 binary frames. The colorbar values range from 50 to 120, which represent the values of the estimated parameter h_n in Equation 3. The higher value represents a higher incident angle.

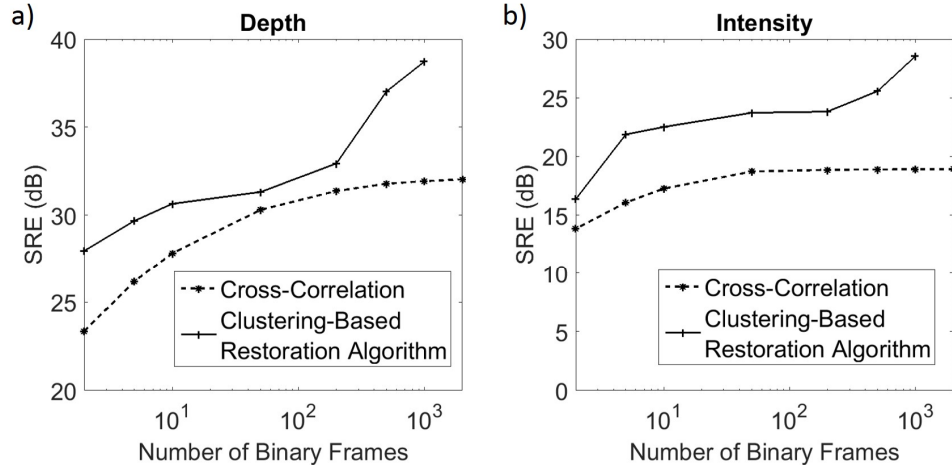


Fig. 13. Signal-to-reconstruction error (SRE) for the polystyrene head target as a function of number of binary frames for both cross-correlation and the clustering-based algorithms. SRE is shown for a) depth profile and b) intensity. SRE for cross-correlation and the median filter corrected cross-correlation algorithms (which show identical SRE values) are represented by the dashed line, and the clustering-based restoration algorithm is represented by the solid line.

intensity profiles when compared to a standard pixel-wise cross-correlation image reconstruction technique, especially in the sparse photon regime. This new method has allowed the accurate estimation of depth images, with millimeter scale features resolved, from as few as two binary frames per delay step, and containing a large number of hot pixels ($>20\%$ of 256×256 pixels). The clustering-based restoration algorithm uses the binomial statistics of the data and non-local spatial correlations within the scene to restore high quality depth estimates surpassing those generated via pixel-wise cross-correlation. As well as depth profiles, this method also provides target's surface local orientation and intensity information. Future work could include the generalization of the clustering-based reconstruction algorithm to improve its robustness to hot pixels and the quality of the estimates.

To the best of our knowledge, this is the first demonstration of a 3D reconstruction of a target scene with millimeter scale depth resolution using a range-gated CMOS SPAD quanta image sensor. The results presented in this paper could point the way towards the practical implementation of target depth profiling using gated photon counting CMOS cameras with large pixel format, high fill-factor and high frame rate without the need for time-correlated single-photon counting functionality. Furthermore, it should be noted that in the measurements shown, the detector range gate was approximately 10 ns and the laser pulse width was as long as 600 ps. The origin of the excellent time resolution is the 10 ps resolution delay. This means that this millimeter resolution depth imager can be used with relatively inexpensive and long duration pulse lasers compared to the short pulse lasers required for high-resolution TCSPC LiDAR systems. Whilst this profiling technique provides potential speed advantages, it is however unlikely to rival the surface-to-surface resolvability (e.g. for target in clutter scenarios) of TCSPC based sensors [38, 39] as the relatively slow rise time will serve as a lower limit.

The image sensor was operated at room temperature which resulted in high dark count rates. Future work could incorporate thermoelectric cooling onto the sensor chip in order to reduce the dark count rate. This would increase the signal-to-noise ratio of the return data and improve the

resolution of reconstructed data recorded at low binary frame numbers. A short exposure time per frame ($0.5 \mu\text{s}$) was used in our experiments, however delay refresh time and data transfer time also affect the total measurement duration and should therefore be kept to a minimum. Future work could integrate an on-board digital delay generator capable of short step size (comparable to that used above) in order to reduce the dead time, thereby increasing the delay iteration rate and allowing for a higher laser repetition rate. Also, frame acquisition and image reconstruction could both be handled by high-performance FPGA boards (e.g. Altera Stratix-V board, which has been used for real-time image processing in [40]) for the camera to minimize the raw frame data transfer time to provide the capability for on-the-fly or embedded image processing [41]. All these approaches to reduce the camera's measurement time would provide a pathway towards real-time range-gated depth profiling. This would make our imaging approach suitable for 3D face and gesture recognition in the fields of surveillance, gaming and augmented reality at indoor ranges, as well as underwater imaging applications [42,43]. For outdoor applications at longer ranges (e.g. with autonomous vehicles) additional filtering approaches would be required to reduce background light and use of higher power laser sources would be necessary to increase the detection SNR. The effect of ambient background illumination on the sensor performance is the subject of future work.

Table 2. Total acquisition time over 1501 delay steps (t_{aq}) for different numbers of binary frames per step, average photon number per pixel over all binary frames (PN_{Σ}) for all delay steps used and total exposure time (t) for cross-correlation (X-Corr.) and clustering-based restoration (Clust.) algorithms

Binary Frames	2000	1000	500	200	50	10	5	2
t_{aq} (s)	759	416	245.5	120.3	87.5	84.1	83.8	83.5
$\text{PN}_{\Sigma, X\text{-Corr.}} (\times 10^3)$	132	65.8	32.9	13.1	3.26	0.65	0.33	0.13
$\text{PN}_{\Sigma, \text{Clust.}} (\times 10^3)$	84.0	42.0	20.9	8.39	2.09	0.42	0.21	0.08
$t_{X\text{-Corr.}}$ (ms)	1501.00	750.50	375.25	150.10	37.53	7.51	3.75	1.50
$t_{\text{Clust.}}$ (ms)	750.00	375.00	187.50	75.00	18.75	3.75	1.88	0.75

Funding

Engineering and Physical Sciences Research Council (EPSRC) (EP/J015180/1, EP/N003446/1, EP/M01326X/1, EP/K015338/1); Royal Academy of Engineering under the Research Fellowship scheme (RF201617/16/31); European Research Council through the EU's Seventh Framework Program (FP/2007-2013)/ERC under Grant 339747.

Acknowledgments

The SPAD array was designed at *The University of Edinburgh* and fabricated at *STMicroelectronics*.